AN EFFICIENT FRAMEWORK FOR SCIENTIFIC ARTICLE REOMMENDATION SYSTEM

Synopsis submitted to CHRIST (Deemed to be University)
as partial requirement for the award of Degree of

DOCTOR OF PHILOSOPHY
IN
COMPUTER SCIENCE

AKHIL M NAIR

1840072

Under the Supervision of
JOSSY P GEORGE
Professor

CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE · INDIA

Signature of the Research Scholar Signature of the Research Supervisor

CENTRE FOR RESEARCH

CHRIST (Deemed to be University)

BENGALURU

OCTOBER 2021

CENTRE FOR RESEARCH

Approval of PhD Synopsis

[as per clause 11(k.) of CU Regulation for the Degree of Doctor of Philosophy, May 2019]

Name of Scholar:	Akhil M Nair
Department:	Computer Science
Names of Research Advisory Committee: 1. Dr. Vivek Deshpande - Director, Vishwakarma Institute of Information Technology, Pune 2. Dr. K. Balachandran - H.O.D , Department of Computer Science and Engineering, CHRIST (Deemed to be University), Kengeri Campus	
Name and Designation of External Reviewer if applicable:	
Date of Synopsis Presentation:	
Title of the PhD thesis:	An Efficient Framework for Scientific Article Recommendation System
Approved by:	
1. Name of Supervisor: Dr. (Fr.) Jossy P. George Signature:	
2. Name of RAC Member: Dr. K. Balachandran Signature:	
3. Name of RAC Member: Dr. Vivek Deshpande. Signature:	

Head of Department

Date of approval by Doctoral Committee:

Director Centre for Research:

Abstract

Excess data makes it challenging to extract information that is relevant to a domain of study or research. Existing state-of-the-art systems focus majorly on the selection of highly connected, prestigious and cited articles, regardless of the relevance of papers. To improve quality of findings, recommender systems which are a subclass of information filtration systems are used. They filter out relevant information over prestigious data from an existing repository of information. There are various sub-domains under recommender systems. This study focuses on citation recommendation. Citations are an integral part of any scientific paper, academic dissertation or projects. Finding appropriate citations for any work is a scholar's most time-consuming task. Thus, a well-defined citation recommendation system provides fulfillment and completeness for citing the giants' works. The thesis aims to study existing frameworks for citation recommendation systems and identify the best dataset to work on graph-based recommender systems. A framework that recommends the most similar and relevant article to the user rather than prestigious authors or papers is here by proposed. The study explores various machine learning and deep learning techniques and methods which can be used effectively in recommending loosely connected yet highly relevant articles.

Keywords: Citation Recommendation System, Scientific Article Recommender, Graph-Based recommendation, Reinforcement Learning, Graph Convolutional Networks, Content-Based Recommendation

TABLE OF CONTENTS

Sr. No.	SECTION TITLES	PAGE No.
1	Abstract	
2	Introduction	1
3	Literature Review and Research Challenges	2
4	Problem Statement and Objectives	6
5	Methodology	7
6	Experiment and Evaluation	10
7	Conclusion and Future Work	25
8	References	26
9	List of Publications (Journals and Conferences)	

TABLE OF FIGURES

FIGURE NUMBER	FIGURE TITLE	PAGE NUMBER
1	Process Flow	7
2	AAN Dataset	8
3	Representation of Data in CORA	9
4	Working Principle of SABED	11
5	The Proposed C-SAR Model	14
6	GRU Similarity Scores	16
7	GRU Adjacency Matrix	16
8	Frequently Cited Documets in AAN	17
9	Phase I of the Proposed Framework	18
10	Score of the DRRC Model	20
11	Workflow of the GCN Analysis	22
12	GCN Classification Report	24
13	Training a validation Loss	24
14	Training and validation accuracy	24

LIST OF TABLES

TABLE NUMBER	TABLE TITLE	PAGE NUMBER
1	Result Comparison of CRS, UBCF and IBCF	12
2	Algorithm of CSAR	15
3	Algorithm of DRRC	19
4	Comparison Of DRRC with State-Of-The Art Models	21

1. INTRODUCTION

In the scientific community, research paper recommendation refers to recommending similar papers to the scholars. The keyword-based search doesn't provide suitable results since the number of items available as results is relatively large. In order to get relevant information from the keyword-based search, the users have to filter out the results. This scientific paper on recommendation system aims at mitigating the information overload. By ranking the publications based on several parameters and recommending the scholars with top-N similar papers based on their area of interest or the target reference, this approach provides researchers with relevant papers. [1]. The recommending systems have grown from a state of supplement to a state of necessity in the academic field. The accuracy and efficiency of the recommender algorithms has improved with time. Recommender systems in recent days are customized from person to person and they are effectively used for a large amount of data over the traditional keyword-based search [2], [3].

There are various types of research article recommendation methods that focus primarily on finding the similarity between articles. They can be classified widely into Content or Context-based Filtering, Collaborative-Filtering, and Hybrid Filtering method. More specifically, these methods can be sub-categorized as: Collaborative [4], Content-based [5], Citation based [6], Meta-data based [7], Multi-level citation network [8], and user-profile based approaches. The Collaborative-Filtering method is mostly used in recommendation systems which focus on discovering the relationship between documents [9], [10]. In e-commerce, Collaborative Filtering treats the cited papers as users and citing papers as items, respectively. The citation network is used to construct a paper-citation matrix. Users are recommended to read the research papers depending on the references generated by different scientific resources. But in case a new article is being chosen as the paper of interest, it will not generate any recommendation until it is cited by other articles. This situation is termed as the cold-start problem. [4]. Content-Based Filtering builds a user profile using the user's historic preferences and personal library. The similarity between the keywords extracted from the user profile and candidate papers is calculated and ranked based on the similarity coefficient. It then recommends the paper with top-n ranks of similarity [5].

Users with similar interests in the past have a probable tendency to retain their similarity in the future. In a Graph-Based method the relationship between the users, papers, and other

parameters is referred to as the edges of the graph and entities like the research articles or papers and the authors are considered the nodes of the graph [8]. A random walk or any such algorithm is used to find the significance between the papers and users by implementing them on these nodes and edges. In the Hybrid Filtering, the above-mentioned methods are clubbed to improve accuracy. Normally Content-Based filtering and Collaborative-Filtering are employed in combination to generate recommendations because they complement each other. Apart from the approaches mentioned above, there is a Latent-Factor model [11] and the Topic-Regression Factorization model [12].

2. LITERATURE REVIEW AND RESEARCH CHALLENGES

In the past twenty years, the number of published studies on research paper recommender systems have crossed two hundred [13]. The authors J. Beel et al., through their study titled "Research-paper recommender systems: a literature survey" discovered that more than half (55%) of the recommender systems for research paper recommendations have used Content-Based Recommendation, 18% used Collaborative-Filtering and 16% have used graph based recommendation methods.

Recommendation systems have existed for decades and they have become a major area of research in the last few years. In this era of big data, users are amazed by the flow of data from numerous sources that are accessible through the internet. They are, however, puzzled about which data to use and how to consume it. They face difficulties in retrieving and processing the most relevant ones when the choices are numerous [14]. The community of scholars has contributed millions of scientific articles till date and the number of papers published are increasing with time. For instance, based on the Digital Bibliography & Library Project (DBLP), scholars and scientists from the computer science field have contributed three times more articles in the current year when compared to the number of articles ten years ago.

With this increasing amount of research articles, the literature search has become a tedious and phenomenal task. On average, a normal research article cites around twenty papers which in turn means the papers that cite or get cited in the reference articles are more than a thousand [15]. For a junior-level researcher, finding a relevant article from the data that is available becomes time-consuming. To tackle this problem, Giles et. al., initiated and presented the concept of a Scientific Article Recommendation System as part of the cite seer project [16]. In the scientific community, research paper recommendations refers to the recommendation of

similar papers to research scholars. This scientific paper on recommendation systems aims at mitigating information overload by providing scholars with relevant papers. It ranks the publications based on several parameters and recommends top-N similar papers based on their area of interest or the target reference [1]. The recommending systems have grown from a state of supplement to a state of necessity in the academic field. The accuracy and efficiency of the recommender algorithms have improved with time. Recommender systems, in recent days, are customized from person to person and are effective for a large amount of data in comparison to the traditional keyword-based search [2] [3]. The keyword-based search doesn't provide suitable results since the number of items available as results are relatively large. The users have to filter out results to get relevant information from the keyword-based search.

In the year 1973, Henry Small in his study formed a new form of coupling, the frequency with which two articles are mentioned together, and defined it as the co-coupling frequency [17]. C. L. Giles, et al., produced a major contribution in the field of research article recommendation with the Cite Seer. It is a self-contained web agent for retrieving and identifying interesting publications. It used the concept of key-word based search and common citations [16]. There were major advancements in the following years which started with a system for personalized tracking of scientific articles [18]. A study lead by R. Dong et al., found that there may or may not exist a social media friendship between different authors. The complete data of the authors also couldn't be included [19]. X. Tang et al., suggested a cross-language context aware recommender system to tackle the challenge of language gap as the citations and contexts are written in various languages [20].

An academic research paper recommender that recommends based on the paper chosen, its main idea was created by Damien Hanyurwimfura et al. [21]. It showed better performance in recommending relevant articles. It also included non-profiled users thereby diminishing the cold-start problem. H. Liu et al., included filtered citation relations in recommending the scientific articles [3]. Dai Hoang Tran et al., proposed a neural network framework using Graph Convolutional Networks (GCN) that require a full adjacency matrix for embedding the learning. The challenge is that the architecture requires a specific format of dataset which the scientific bibliographic datasets cannot convert to. The accuracy measure was also not equal to the state-of-the art baselines [22]. Scientific Paper Recommendation Using Citation Context was proposed by N. Sakib et al., which followed the concept that if two papers are citing a

paper together or if two papers are being cited by a paper together, then the underlying concept would be similar to some extent. This work produced a concept of co-occurrence and citation similarity [10]. C. Jeong et al., modified the model of citation context using the Graph-Convolutional Networks (GCN) layer and Bidirectional Encoder Representations from Transformers (BERT), thereby creating a start for the deep learning techniques. A major challenge of the framework, however, is that the architecture works on a specific dataset that contains the context sentences along with the cited references which do not apply to the available databases [23].

Using a model that took into account the paper's topic and concepts, the primary topics were generated as subqueries. The model helped users to find a group of relevant documents [21]. Based on the NDCG metric, the model achieved an accuracy of 80%, demonstrating that it is one of the most effective methods for making suggestions without a user profile. By indexing characteristics, the model's performance improved. Abdul Samad et al. [24] suggested a centrality-based technique for assessing textual and topological similarity metrics. On the basis of the performance of the Cosine and Jaccard metrics, a comparison was made. When examined on the dataset, topological-based similarity via Cosine was shown to have an accuracy of 85.2 percent, whereas Jaccard had a 61.9 percent accuracy. For a DeepWalk-based technique that employed the deep learning-based bibliographic content-specific recommendations, a recall of 0.2285 and NDCG of 0.3602 were obtained [25]. The model's matrix is based solely on the vector matrix of the papers, and does not depend on the citation information. Only if the article contents and citations are included, will the deep walk-based method be predicted to beat other existing models. A Recurrent Neural Network was used to model a recommendation system in a deep learning-based study [26]. Users' explicit and implicit input, as well as a semantic representation of the paper's title and abstract, were utilized.

2.1 Research Gap

- Deep Learning based graph methods have not been experimented on the generalized bibliographic citation recommendations.
- In most of the work which is done on Deep Learning Framework for citation recommendation, the data set is specially generated in a controlled environment, but the documents which are available in various libraries are in real-time available data format.

- This research aims at implementing Deep Learning-Based models as well as the hybrid approach which majorly focuses on taking advantage of the graphs and their adjacency matrices.

2.2 Challenges

- The size of the dataset is huge and it would require high computational power to build a model using graphical nodes.
- Generalizing the framework for the available data format is a challenging task as the architecture needs to be tuned to the available data and it might reduce efficiency.

3. PROBLEM STATEMENT AND OBJECTIVES

The field of research and innovations has been a major contributor to societal growth for a long time. A major part of a research study is finding relevant studies and research articles [27]. Finding relevant articles over prestigious ones in a domain is a tedious task with the existing techniques of keyword search and other search engines. The research article recommenders should focus on the relevance and the content of the papers than the citations and the authors of the paper to provide an impartial recommendation to scholars.

However, the current state-of-the-art engines are just focused on the number of citations and keyword similarity, to recommend the papers. The research scholars thus spend a huge amount of time searching the papers than studying [28].

Primary Objective:

To design an efficient framework for recommending citations for scientific articles using different query methods.

Secondary Objectives:

1. To improve upon the citation ranking accuracy of the current recommendation system using various learning concepts and techniques.

2. To study the importance of co-citation in analyzing the similarity of various scientific articles.

3. To analyze different recommendation techniques and query methods to compare and analyze the efficiency on a different dataset [AAN and Cora].

4. METHODOLOGY

The proposed framework for the scientific article recommender systems is implemented by using independent experimental models. The experiments were carried out with different learning techniques ranging from basic probability calculation to deep reinforcement learning techniques.

- The first set of experiments (2 models – SABED and CSAR) involves evaluation of the performance of Content-Based Filtering techniques on the datasets and thereby improving the efficiency of the recommendation.
- The second set of experiments (2 models – DRRC and GCN) completely focused on building deep learning frameworks using graph-based learning.

The major steps involved in creating the framework are: Pre-processing, Content-based and Graph-Based filtering, and Evaluation. The process flow diagram for our Recommender Framework is as seen in Fig.1.

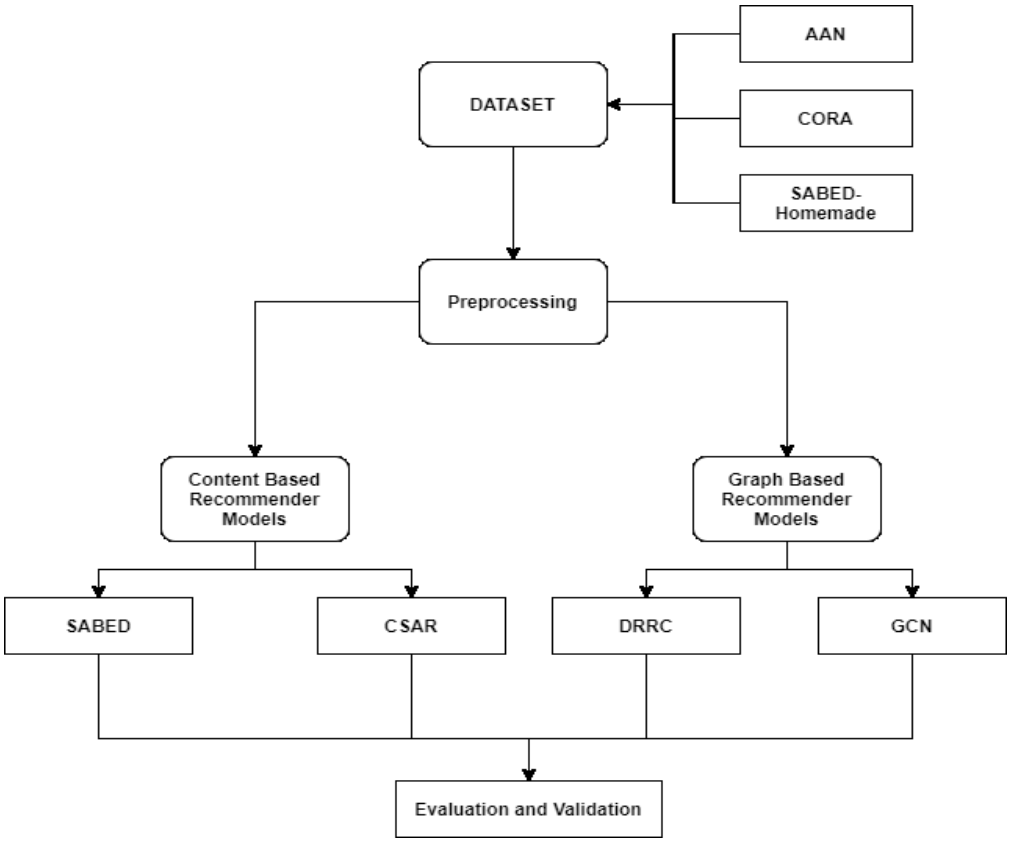


Fig. 1 Process Flow

The experiment to create a new framework started with identifying a dataset that suits the generalized digital library. The identified datasets were ACL Anthology Network (AAN), Cora, and a minor homemade dataset (SABED) with over 70 PDF documents of the articles from the image processing domain. The main focus of the study after dataset identification was to shortlist the preprocessing techniques that are commonly used for content-based filtering and Hybrid Filtering techniques. Content-based recommendation techniques were evaluated with an experimental framework – Similarity Analysis for Citation Recommendation System using Binary Encoded Data (SABED) and Content-Based Scientific Article Recommendation System Using Deep Learning Technique (CSAR). The deep learning frameworks proved to be much more efficient than the content-based framework in terms of scalability and relevance and hence the experiments were carried out using reinforcement learning (DRRC) and Graph Convolutional Networks (GCN). The evaluation of the resulted output was done using the metrics, Accuracy, Recall, and manual cross-verification.

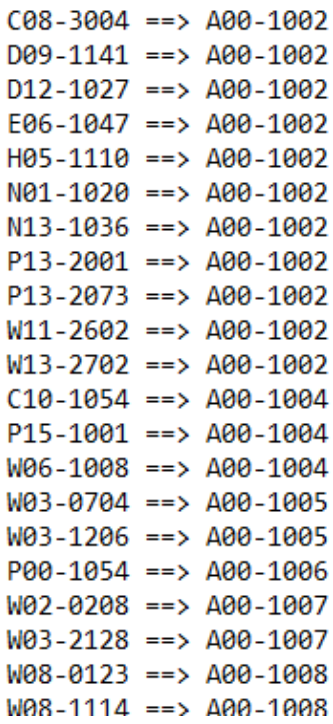
4.1. Description of the Dataset

Scientific recommender systems have grown in popularity due to their ability to overcome information overload by providing researchers with timely and relevant content [29]. Elsevier, PubMed, SpringerLink are some of the leading digital libraries that send alerts or messages about related scientific journals that match user records and interests. These filtering systems that match the user preferences are termed the Scientific Article Recommender Systems. For this study, three different datasets depending on the learning models used. The names of these datasets are ACL Anthology Network (AAN), CORA, and SABED which is a handmade Dataset.

ACL Anthology Network (AAN)

ACL Anthology Network (AAN) [30] is a well-known collection of natural language processing and computational-linguistics articles. AAN is a compilation of Association of Computer Linguistics (ACL's) documents. It includes a variety of networks, that includes citations of paper, author, and author collaboration. The dataset for the suggested study is AAN 2014 published dataset, which is a 759 MB file. The folder contains several text documents, including author citation network, releases, papers text, author affiliations etc. It contains the paper ids of 23,766 publications as well as the papers it cites, totaling 1,24,857 paper citations. 2014 release folder has been used for the study. AAN has the author to paper

citations, releases, metadata of the papers and its ids as shown in th Fig. 2.

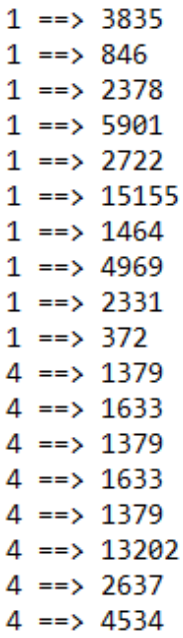


```
C08-3004 ==> A00-1002
D09-1141 ==> A00-1002
D12-1027 ==> A00-1002
E06-1047 ==> A00-1002
H05-1110 ==> A00-1002
N01-1020 ==> A00-1002
N13-1036 ==> A00-1002
P13-2001 ==> A00-1002
P13-2073 ==> A00-1002
W11-2602 ==> A00-1002
W13-2702 ==> A00-1002
C10-1054 ==> A00-1004
P15-1001 ==> A00-1004
W06-1008 ==> A00-1004
W03-0704 ==> A00-1005
W03-1206 ==> A00-1005
P00-1054 ==> A00-1006
W02-0208 ==> A00-1007
W03-2128 ==> A00-1007
W08-0123 ==> A00-1008
W08-1114 ==> A00-1008
```

Paper Citations

```
id = {E03-1001}
author = {Oard, Douglas W.}
title = {Multilingual Access To Large Spoken Archives (Invited Talk)}
venue = {EACL}
year = {2003}

id = {E03-1002}
author = {Henderson, James B.}
title = {Neural Network Probability Estimation For Broad Coverage Parsing}
venue = {EACL}
year = {2003}

id = {E03-1003}
author = {Burstein, Jill; Wolska, Magdalena}
title = {Toward Evaluation Of Writing Style: Overly Repetitious Word Use}
venue = {EACL}
year = {2003}

id = {E03-1004}
author = {Cmejrek, Martin; Curin, Jan; Havelka, Jiri}
title = {Czech-English Dependency Tree-Based Machine Translation}
venue = {EACL}
year = {2003}

id = {E03-1005}
author = {Bod, Rens}
title = {An Efficient Implementation Of A New DOP Model}
venue = {EACL}
year = {2003}
```

Metadata



```
1 ==> 3835
1 ==> 846
1 ==> 2378
1 ==> 5901
1 ==> 2722
1 ==> 15155
1 ==> 1464
1 ==> 4969
1 ==> 2331
1 ==> 372
4 ==> 1379
4 ==> 1633
4 ==> 1379
4 ==> 1633
4 ==> 1379
4 ==> 13202
4 ==> 2637
4 ==> 4534
```

Author to paper citations

```
W09-2307    Discriminative Reordering with Chinese Grammatical Relations Features   2009
W04-2607    Non-Classical Lexical Semantic Relations    2004
W01-1314    A System For Extraction Of Temporal Expressions From French Texts Based On Syntactic And Semantic Constraints   2001
W04-1910    Bootstrapping Parallel Treebanks    2004
W09-3306    Evaluating a Statistical CCG Parser on Wikipedia    2009
W91-0219    Development Of The Concept Dictionary Implementation Of Lexical Knowledge    1993
W97-0806    Integrating A Lexical Database And A Training Collection For Text Categorization    1997
W98-1121    POS Tagging Versus Classes In Language Modeling 1998
W11-0819    FipsCoView: On-line Visualisation of Collocations Extracted from Multilingual Parallel Corpora  2011
W00-1221    Zero Anaphors In Chinese Discourse Processing   2000
W02-0506    Building A Shallow Arabic Morphological Analyser In One Day     2002
W04-1215    Annotating Multiple Types Of Biomedical Entities: A Single Word Classification Approach 2004
W11-2152    Two-step translation with grammatical post-processing   2011
W09-2401    Invited Talk: Alternative Annotations of Word Usage     2009
W07-2037    I2R: Three Systems for Word Sense Discrimination Chinese Word Sense Disambiguation and English Word Sense Disambiguation    2007
W07-0719    Context-aware Discriminative Phrase Selection for Statistical Machine Translation    2007
W03-1811    A Disambiguation Method For Japanese Compound Verbs     2003
W98-0314    Signalling In Written Text: A Corpus-Based Approach     1998
W11-2102    A Lightweight Evaluation Framework for Machine Translation Reordering   2011
W09-0109    What Do Computational Linguists Need to Know about Linguistics? 2009
W04-1709    Sentence Completion Tests For Training And Assessment In A Computational Linguistics Curriculum 2004
H89-2009    Answers And Questions: Processing Messages And Queries  1989
W98-1427    Generation As A Solution To Its Own Problem     1998
W04-3112    Using Natural Language Processing LocusLink And The Gene Ontology To Compare OMIM To MEDLINE    2004
W98-0505    Types Of Syntagmatic Grammatical Relations And Their Representation     1998
W05-1516    Strictly Lexical Dependency Parsing     2005
W08-0204    Defining a Core Body of Knowledge for the Introductory Computational Linguistics Curriculum     2008
W00-0304    NJFun - A Reinforcement Learning Spoken Dialogue System 2000
W98-1217    Cross-Entropy And Linguistic Typology   1998
W00-1103    Use Of Dependency Tree Structures For The Microcontext Extraction       2000
W94-0310    On The Creative Use Of Language: The Form Of Lexical Resources  1994
```

Paper Titles with the IDs

Fig. 2 AAN Dataset

CORA - Relational Dataset Repository

The Cora dataset contains 2708 scientific publications which are divided into seven categories. Machine Learning papers make up the Cora dataset. These papers are categorized into one of seven categories:

a. Case_Based

b. Neural_Networks

c. Probabilistic_Methods

d. Reinforcement_Learning

e. Rule_Learning

f. Theory

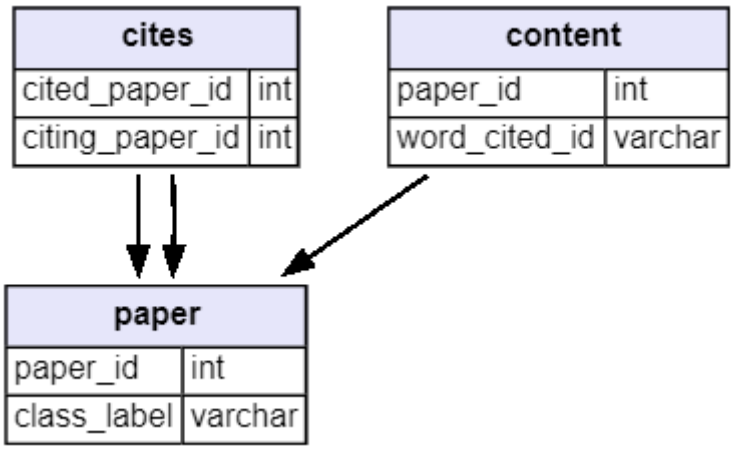


Fig. 3 Representation of Data in CORA

The papers were chosen in such a way that each one cites or is cited by at least one other paper in the final corpus. The total number of papers in the corpus is 2708. After stemming and removing stop words, we discovered a vocabulary of 1433 distinct words. The list was culled of all words having a document frequency of less than ten. There are 5429 linkages in the citation network. A 0/1-valued word vector describes each publication in the dataset and indicates the existence or absence of the associated word from the dictionary. There are 1433 distinct terms in the dictionary. The graphical representation of the data available in the dataset is shown in Fig 3.

5. EXPERIMENTS AND EVALUATION

5.1. Similarity Analysis for Citation Recommendation System using Binary Encoded Data (SABED)

One of the key difficulties of the scientific article recommendation system using citations is establishing the users' preferences when the user profile is new or new things are uploaded to the database and this state is known as cold start condition [25]. Furthermore, when two individual words have the same connotation or when few words are insufficient to describe the context of work, the keyword-based similarity problem develops [31][32]. The search result of keyword-based similarity is not efficient enough to detect the important information in these circumstances. When there are self-citations and manually adjusted citation counts, the recommendations become skewed.

However, the study SABED, examines the similarity between various publications by converting the text metadata from the article into a binary encoded format, such as the authors' names, the paper's DOI, keywords, abstract, and content [33]. Thus, by considering the search probability, it is simple to determine the similarity, resulting in improved accuracy[34].

SABED is a binary encoded similarity analysis of data that uses search probability. It calculates the exact match between the documents and hence determines the degree of similarity. SABED uses ASCII values to transform the Abstract,Keywords, Author_Name and to Binary values of 0 and 1. The proposed citation recommendation's working concept is depicted in Fig. 4, which starts with the database collection for the scientific publication and concludes with a prediction and recommendation for the specified database query. The SABED model makes use of a custom dataset of more than 70 PDF documents from the Image Processing domain.

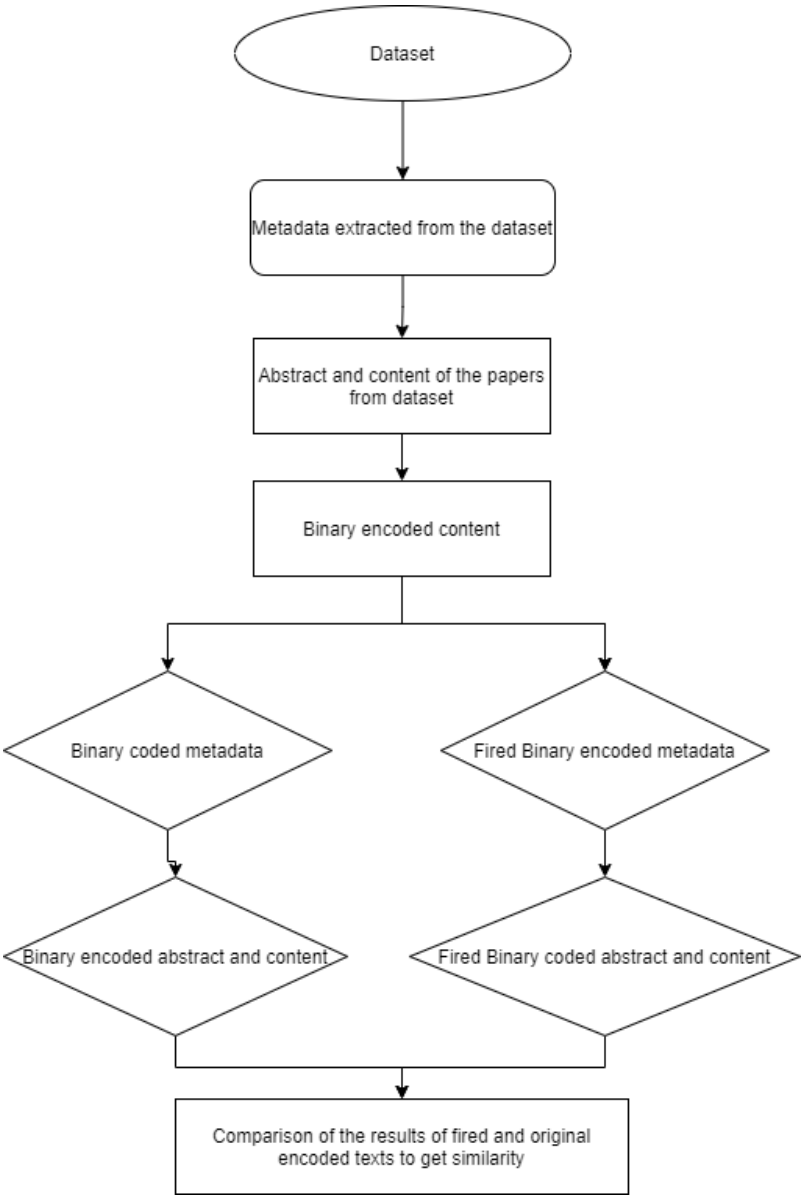

Dataset
Metadata extracted from the dataset
Abstract and content of the papers from dataset
Binary encoded content
Binary coded metadata
Fired Binary encoded metadata
Binary encoded abstract and content
Fired Binary coded abstract and content
Comparison of the results of fired and original encoded texts to get similarity

Fig 4. Working Principle of SABED

To develop the data for the suggested framework, a large array of scientific research papers were gathered, exclusively from Willy, ACM, IEEE digital library and Springer. SABED, translates text metadata from an article into binary format. The authors' names, the publication's DOI, keywords, the abstract, and the substance of the paper can be converted. In a binary encoded format, the query is sent to the database [35]. As a result, when a query containing the binary data format matches any of the parameters from the database, such as the author's name, keywords, DOI, abstract, and content of the research articles, the search probability is displayed, reflecting the degree of similarity between the fired query and the knowledge base contents.

Result of SABED:

Table 1. Result Comparison of CRS, UBCF, and IBCF

Algorithm	*Accuracy 1*	*Accuracy 2*	*Accuracy 3*
CRS	0.95	0.70	0.47
UBCF	0.86	0.5	0.35
IBCF	0.68	0.545	0.4

One of the evaluation metrics utilized to compare the proposed technique to IBCF and UBCF is accuracy. Equation (1) shows how to calculate the algorithm's accuracy using a formula. The number of right predictions made as a percentage of the total number of predictions generated by the algorithm is used to determine accuracy. Table 1 compares SABED, IBCF, and UBCF, three different types of recommender systems.

$$Accuracy = \frac{Recommendation\ obtained\ using\ the\ probability}{Total\ recommendations\ obtained} \quad \ldots (1)$$

5.2. Content-Based Scientific Article Recommendation System Using Deep Learning Technique (CSAR)

Scientific users can now obtain scholarly articles from a pool of available papers with greater efficiency and precision because of the big data era. An information overload problem has developed as the publications happening per year is increasing at an exponential rate, making it impossible for researchers to perform thorough literature reviews. This difficulty can be solved with the use of a scientific article recommendation system that provides the users with customized and personalized recommendations based on their interests and preferences. Two commonly used approaches for recommendations are Collaborative Filtering (CF) and Content-Based Filtering (CBF). Even though article recommendation systems have progressed significantly, the development of a content-based deep learning algorithms are still in its infancy [36]. In this paper, a CSAR model is suggested that uses the Gated Recurrent Unit (GRU) and Apriori algorithm to generate article recommendations based on content similarity. In terms of suggesting related publications, the combination of a deep learning based technique and a classical algorithm in data mining is projected to outperform the current state-of-the-art model.

A Content Based Scientific Article Recommendation (C-SAR), which integrates the advantages of both GRU and Apriori algorithms to locate related sets of articles is proposed in this study. The most similar collection of texts will be filtered using a GRU and an Apriori hybrid approach.

Gated Recurrent Unit (GRU)

In the Recurrent Neural Network, GRU is a gating approach. This architecture is designed to handle large data sequences. GRU is an improved form of traditional Recurrent Neural Networks that uses two gates- an update gate and a reset gate - to reduce the vanishing gradient problem of a conventional RNN. These gates act as two vectors that determine which data should be sent to the output. The gating technique establishes controil on the memory of the values that are processed throughout time. This is made up of two gates that govern data flow via states which are update gate and the reset gate. These gates can be thought of as vector entries that combine in a convex way. When necessary, they determine which hidden state should be reset or updated. As a result, the network learn to filter the transitory observations that are not relevent.

Apriori Algorithm

As a rule-based machine learning technique, association rule learning is applied to find hidden patterns and associations between variables in huge databases. Aprıori is a frequent occurring pattern mining algorithm. It finds the most common individual itemsets and expands them to larger sets of items till they appear in the database frequently. Support, confidence, and lift are the three most popular measures to measure relationships. By tracking the equated fractions of transactions in which an item set appears, support determines the item popularity with its number of occurences. Confidence determines the likelihood of one item being procured if another item is bought. Lift is comparable to confidence in that it determines the likelihood of an item (Y) being procured when another item (X) is bought while keeping Y's popularity in check. This algorithm's core characteristic asserts that all subsets of a frequently occuring itemset should also be frequent, and that if an itemset is infrequent, so will all of its supersets.

Model Description

The C-SAR model's technique is divided into two phases as in Fig.5. The GRU technique is employed in the first phase to determine both document similarity and the adjacency matrix.

The Apriori method is used in the second phase to select the most relevant collection of papers from a pool of similar documents.

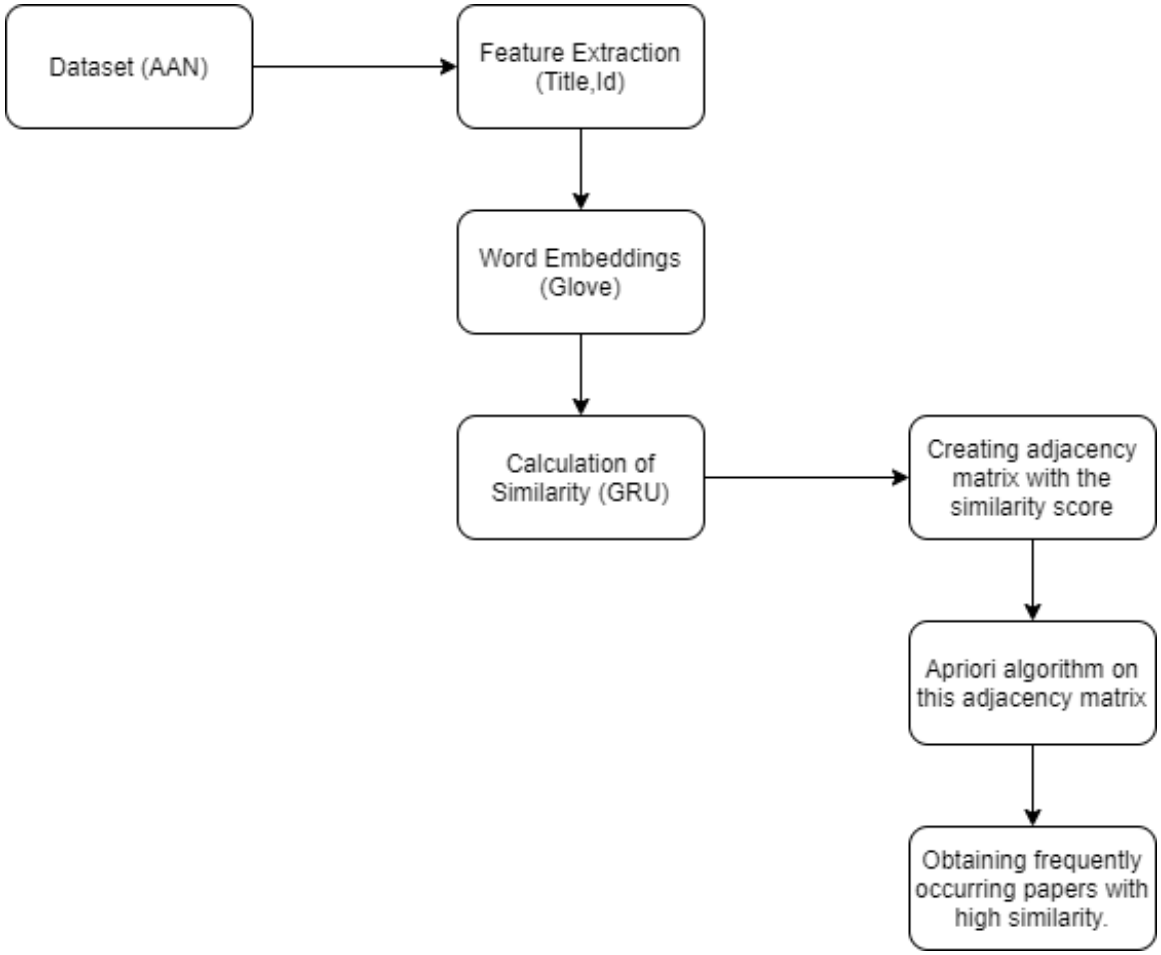

Dataset (AAN)
Feature Extraction (Title,Id)
Word Embeddings (Glove)
Calculation of Similarity (GRU)
Creating adjacency matrix with the similarity score
Apriori algorithm on this adjacency matrix
Obtaining frequently occurring papers with high similarity.

Fig. 5 The Proposed C-SAR Model

The ‘title' feature is taken from the AAN dataset, which is then cleaned as the titles will have null and other irrelevant words. The data has been cleansed of null values, missing values or insignificant values. The ‘title' feature was improved by removing stop words, stemming, and lemmatization for better results. After one-hot encoding, to avoid ambiguity, the text data is padded to ensure that the sequences of data are all of the same length. It is used to ensure that all sequences in a batch are of the same size. The algorithm of the CSAR model is as seen in the Table 2.

Table 2. Algorithm of CSAR

Step 1	The feature (title) is extracted from the AAN dataset.
Step 2	The text data is converted to vectors using Glove pre-trained embedding.
Step 3	Similarity probability is calculated for the documents using GRU
Step 4	The probabilities are replaced by 1's and 0's based on a fixed threshold
Step 5	An adjacency matrix is created with the new values(1's and 0's)
Step 6	The new matrix with similarity score is passed to the Apriori Algorithm
Step 7	The set of frequently occurring documents are obtained

GloVe (short for "Global Vectors") is an unsupervised learning method for obtaining word vector representations. This is achieved by word mapping them into a meaningful space with other comparable words. They are well trained on large datasets, because of which they capture syntactic and semantic information using such pretrained models. They also help the model perform better. Glove's main principle is to extract associations between words, and it has proven to be more effective than other embedding methods like Word2Vec. The likelihood of similarity between input sequences is calculated by GRU units using a vector representation. The similarity score was determined by passing two sequences to GRU model at the same time. Filtering the most and least comparable documents was done using a 0.55 threshold. Based on the likelihood scores, the results are modified by replacing them with 1's and 0's. The presence of a 1 denotes that the papers are similar, but a 0 suggests that they are not. The results are turned into an adjacency matrix.

The adjacency matrix is then passed on to the Apriori algorithm to identify the frequent occurences of documents from the AAN dataset. For filtering out the most similar and frequently occurring documents, the framework employs a dual-layered filter. The first layer uses the GRU algorithm to calculate document similarity, while the second layer uses the GRU output of 0s and 1s as a paper-paper adjacency matrix. The support is set to 0.7, and confidence

levels are calculated using the frequency of the event.

With the integration of GRU, and association rule mining, the proposed C-SAR model is projected to outperform existing state-of-the-art methods. The proposed method's objective is to identify the set of documents that are the most comparable from the AAN dataset. The use of GRU in C-SAR model produced the probability score of similar documents in the first phase. GRU generated the similarity score for the paper-paper matrix as shown in Fig. 6.

p1_encoded	p2_encoded	predictions_probs
[138, 174, 18, 28, 90, 51, 64]	[105, 116, 31, 15, 51]	0.504709
[138, 174, 18, 28, 90, 51, 64]	[4, 11, 2, 22, 1, 139, 117, 19, 175, 10, 32, 1...	0.588963
[138, 174, 18, 28, 90, 51, 64]	[140, 79, 107]	0.504478
[138, 174, 18, 28, 90, 51, 64]	[25, 4, 43, 309, 141, 12, 176]	0.539213
[138, 174, 18, 28, 90, 51, 64]	[495, 12, 312, 496, 1, 497, 22, 19, 71, 79, 45]	0.544077

Fig. 6 GRU Similarity Scores

Fig. 5 shows the adjacency matrix created by replacing the probabilities obtained from the GRU with 1s and 0s according to the threshold value. To produce a cleaner data for the association rule mining-based model, the adjacency matrix is created. The Apriori algorithm employs this adjacency matrix to find the most common paper titles and ids in the collection, narrowing down the number of relevant publications.

	W09-2307	W04-2607	W01-1314	W04-1910	W09-3306	W91-0219	W97-0806	W98-1121	W11-0819	W00-1221
W09-2307	1.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0
W04-2607	1.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0
W01-1314	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0	1.0
W04-1910	1.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0
W09-3306	0.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	1.0	1.0

Fig. 7 GRU Adjacency Matrix

The Apriori algorithm was fed the adjacency matrix displayed in Fig 7. The algorithm's support value is 0.35, which was used to eliminate all data items with a frequency of less than 35

percent, yielding the findings given in Fig 8.

	support	itemsets
0	0.848	(W09-0109)
1	0.714	(W98-1405)
2	0.878	(W00-0502)
3	0.702	(W04-1406)
4	0.816	(H92-1028)
...	...	...
2914	0.600	(W97-0616, W00-0502, W09-0109, W03-0305, W04-1...
2915	0.604	(W00-0502, W09-0109, W03-0305, W04-1406, H92-1...
2916	0.622	(W97-0616, W00-0502, W04-0505, W09-0109, W07-0...
2917	0.606	(W97-0616, W00-0502, W09-0109, W07-0201, W03-0...
2918	0.624	(W97-0616, W00-0502, W04-0505, W09-0109, W03-0...

Fig. 8 Frequently Cited Documents in AAN

After analyzing the data, a minimum of seven occurrences of the same itemset was evaluated. A deep learning technique was used to develop the CSAR model. This technique focuses on comparing papers based on how similar their titles are. For document similarity, a Gated Recurrent Unit approach was used, and association rules were used to select the frequently occurring set of documents from a similar set. This novel approach outperforms current methods that leverage user representations and simple K-Means Clustering.

This model's limitations are the time and memory limits during the implementation phase. Using cloud servers and other super configured machines we can boost performance. The model's efficiency can be improved in the future by taking into account an optimal threshold value for obtaining the similarity matrix and a better minimum support count based on the total number of transactions.

5.3. Citation Recommendation System using Deep Reinforcement Learning

Deep Reinforcement Recommender for Citations (DRRC) is a proposed framework that employs reinforcement learning technique to learn the available paper-by-paper-citation network to produce the most appropriate recommendations. Fig. 9 depicts the DRRC model that has been proposed.

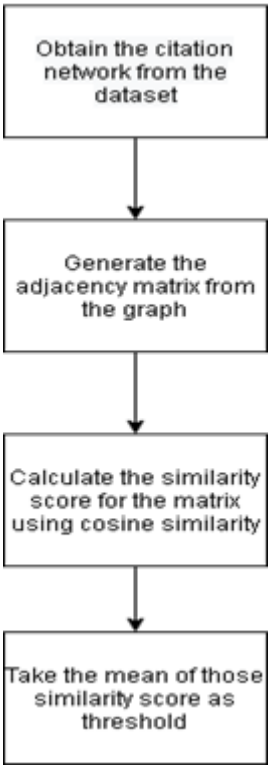

Obtain the citation network from the dataset
Generate the adjacency matrix from the graph
Calculate the similarity score for the matrix using cosine similarity
Take the mean of those similarity score as threshold

Fig 9. Phase one of the Proposed Framework

Model Description

The proposed model comprises two parts: generating the threshold for the algorithm, and applying reinforcement learning to the citation network using the acquired threshold.

The model initially creates the corresponding adjacency matrix from the citation of paper-by-paper network to determine the threshold for the algorithm. The cosine similarity method is then applied to calculate the matrix's similarity values. The mean of the similarity matrix is used as a threshold for the algorithm once it has been retrieved.

The state, action, and reward components make up the algorithm of the reinforcement learning framework. The calculation of the similarity value is the model's current condition. The action helps to compare the threshold and similarity values. The model receives a positive reward 1 instead of a negative reward 0 if the obtained score of similarity exceeds the threshold. A new adjacency matrix is generated based on these incentives, which has a higher number of links than the prior one. The similarity score for the newly generated matrix is then calculated once more. This technique is repeated for 1,00,000 iterations, with each iteration selecting two nodes

at random that have never been joined before. At each iteration of the model, the mean of the similarity matrix is determined, revealing the model's learning output. The model's workflow is depicted in Table 3. Finally, the model's Recall@N value is calculated, where N = [20, 40, 60, 80, 100].

Table 3. Algorithm for DRRC

Step 1	Create the citation network graph from the formatted dataset.
Step 2	Create the adjacency matrix of the graph obtained.
Step 3	Obtain the similarity matrix which consists of the similarity value between two nodes. The similarity values are calculated with the help of cosine similarity from the adjacency matrix.
Step 4	Take the mean of the similarity values as the threshold.
Step 5	Randomly select 2 unconnected nodes.
Step 6	If the similarity value in Step 5 is above the obtained threshold, change the corresponding value in the adjacency matrix at Step 2 to 1 else to 0.
Step 7	Recalculate the mean of cosine similarity of the new adjacency matrix.
Step 8	Keep repeating Step 5 to 7 for 1 lakh iterations.
Step 9	Plot all the newly obtained similarity values after every iteration of the model for evaluation.
Step 10	Check Recall@N values for the newly obtained matrix.

Experimental Result of DRRC:

The DRRC model is scored and validated by calculating and visualising the mean similarity value of the adjacency matrix at each iteration. This will reveal the model's learning rate as well as the quality of the similarity it generates after each iteration.

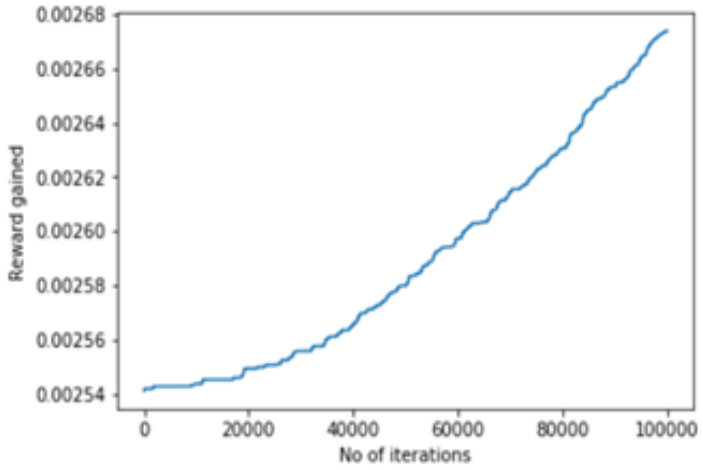


Fig. 10 Score of the DRRC Model

Before 20000 iterations, the model struggled to extract the nodes that should be connected, as shown in Fig. 10, but after the 20000 iterations, there is a rapid increase in the reward, suggesting that it has learned the positives and negatives from the previous calculations.There are various performance metrics available for the evaluation of a recommender system such as –

a) For some given N, Recall@N determines the mean of the percentage of papers from the test set that appear in the top N ranked list.

b) The average precision score for each text in a set of manuscripts is called Mean Average Precision (MAP).

c) When there are many levels of relevant evaluation, the Normalized Discounted Cumulative Gain (NDCG) quantifies the correctness of rankings.

The DRRC model's performance metric is Recall@N, with N values of 20, 40, 60, 80, and 100. Table 4 compares the proposed DRRC model's metric values to those of other state-of-the-art models. Our DRRC model outperforms state-of-the-art models for Recall@N values, as can be observed.

Table 4. Comparison of DRRC with State-of-The Art Models

Model Name	Recall@20	Recall@40	Recall@60	Recall@80	Recall@100
DRRC	0.552	0.591	0.619	0.641	0.655
QR	0.242	0.331	0.401	0.425	0.479
CQR	0.239	0.328	0.396	0.421	0.475
QR-PV	0.219	0.285	0.348	0.367	0.411
QR-PA	0.215	0.281	0.343	0.362	0.402
QR-P	0.206	0.273	0.339	0.351	0.396
CQR-P	0.206	0.269	0.332	0.349	0.395
Cosine	0.196	0.253	0.319	0.337	0.382
CF	0.196	0.250	0.318	0.335	0.379
Okapi	0.193	0.250	0.315	0.334	0.377
K-L	0.191	0.249	0.312	0.331	0.376

5.4. A Graph Convolutional Network based Document Classification

This study investigates the accuracy and efficiency of the Graph Convolutional Networks (GCN) on the Cora Dataset in classifying articles based on citations and article class. Its goal is to

demonstrate that GCN-based networks can classify articles with exceptional accuracy.

Model Description:

There are 2708 nodes in the Cora dataset utilized in this implementation. A scientific paper is represented by each node in the citation network. The node features are represented as a bag of words with a binary value of "0" or "1," indicating whether or not the word is present in the line.

The dataset is a unidirectional graph that displays whether one document cites another or vice versa. After the data has been loaded, a node matrix X is created, as well as a list of tuples containing the nearby nodes. The list of tuples with nearby nodes is then used to generate an adjacency matrix A. There are 1433 words in the node features.

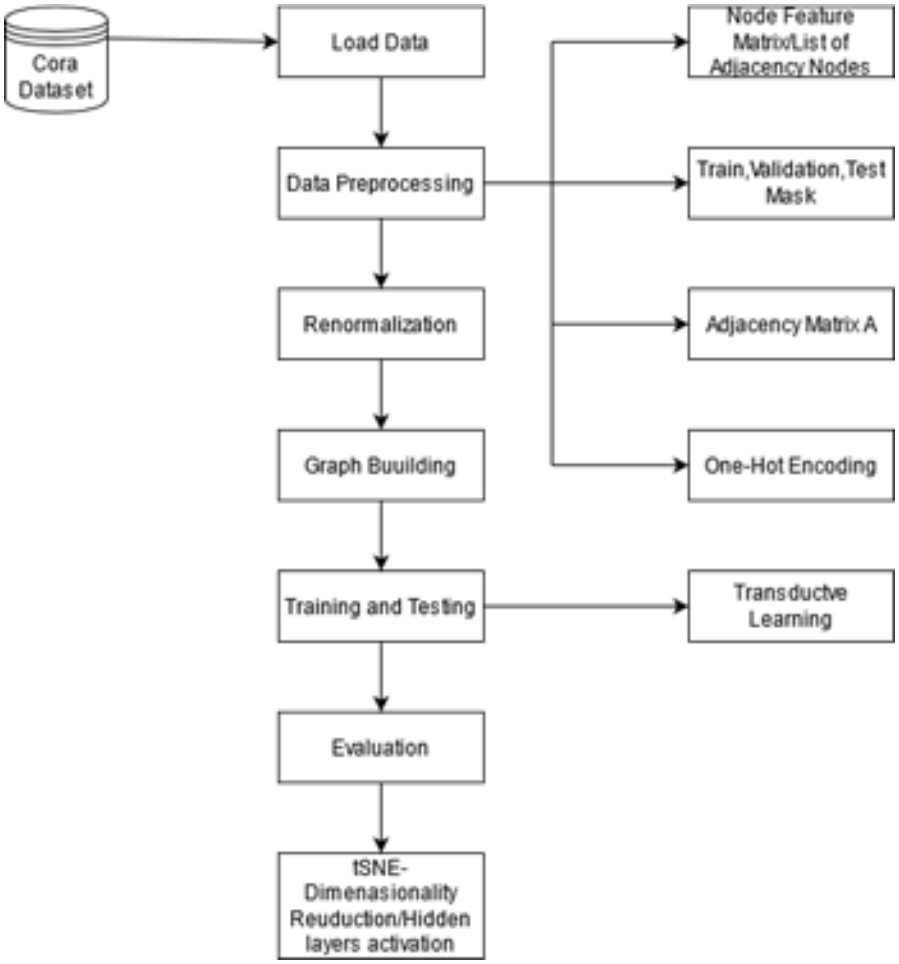


Fig.11 Workflow of the GCN Analysis

A test, train, and validation mask are created during the preprocessing step so that the associated data is available to the corresponding phase, as shown in Fig. 11. The adjacency

matrix A is generated as the next step in the preprocessing. This adjacency matrix A is used to create a node and edge graph. With an average degree of 3.89, graph G contains 2708 and 5728 edges. One-hot encoding is used to convert the category labels to numerical categories ranging from 0 to 7. The node feature matrix N and the adjacency matrix A will be used as inputs to the model. The experiment focuses on a two-layer GCN with elastic regularisation and dropout layers. The first layer employs ReLU activation, whereas the second employs the Softmax technique. The model is built with an epoch of 200 and an early stopping with a patience value of 10.

Before the preprocessed data is transferred to the model, it must go through renormalization, which employs an identity matrix to add the self-loops for the adjacency matrix.

Experimental Results of GCN

The GCN is shown to be efficient enough to use on structured data with classification labels. The efficiency of each layer is also affected by the regularisation techniques used. Methods of activation such as ReLU and Softmax are also important in enhancing efficiency. The model's layer count is limited to two. F1-Score is utilized instead of accuracy and loss measurements to evaluate the model effectively. The proposed method can be applied to any structured data. On the f1-score, the model had an accuracy of 0.75 or 75%. As shown in Fig. 12, the model's macro average f1-score value is 0.74.

GCN Classification Report:

	precision	recall	f1-score	support
Case_Based	0.75	0.69	0.72	114
Genetic_Algorithms	0.90	0.85	0.87	156
Neural_Networks	0.82	0.70	0.76	290
Probabilistic_Methods	0.80	0.76	0.78	172
Reinforcement_Learning	0.70	0.82	0.76	85
Rule_Learning	0.55	0.77	0.64	60
Theory	0.56	0.70	0.62	123
accuracy			0.75	1000
macro avg	0.73	0.76	0.74	1000
weighted avg	0.76	0.75	0.75	1000

Fig. 12 GCN Classification Report

When semi-supervised learning and supervised learning models are used, the GCN-based

framework is very efficient when working with structured data. The accuracy and loss of epoch training have been represented in Fig. 12. The graph shows that the model minimizes the loss value and improves the accuracy value on validation data with each epoch.

As can be shown in Figs. 13 and 14, the model minimises the loss value and improves the accuracy value on validation data with each epoch. The GCN model improves training accuracy by taking into account not just the connections between nodes but also the node attributes.

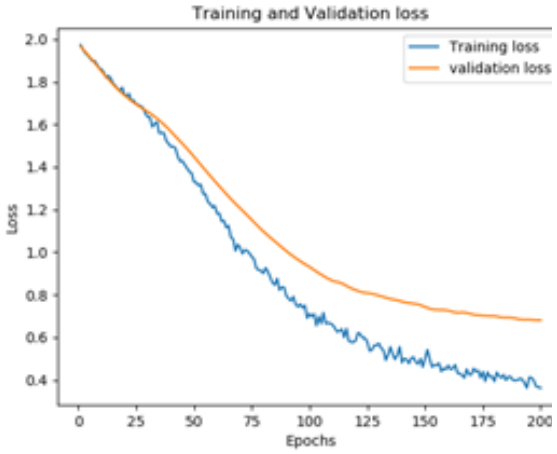


Fig. 13 Training a validation Loss

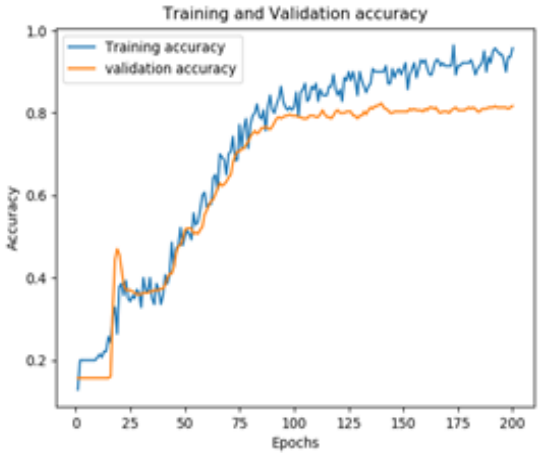


Fig. 14 Training and validation accuracy

6. CONCLUSION AND FUTURE WORK

Recommender systems are one of the most commercially researched topics for business development. Recommender techniques can be categorized as collaborative filtering, content-based filtering, graph-based recommenders, and hybrid recommenders. Even though the domain of recommender systems has been widely researched, citation recommender systems need major modifications in terms of scalability and relevance. Currently, available citation recommendation systems focus majorly on prestigious citations over relevant ones. The major challenge in the domain of citation recommender systems is the sparsity in data. Diversity in the research domain of an individual author is also a challenge. Hence, state-of-the-art collaborative and content-based recommender systems use implicit and explicit feedback but focus on a single type of domain. Another challenge with the citation is scalability. The study proposes a framework using deep reinforcement learning to improve the ranking accuracy and relevance of the citations. The results from the framework are used to build a prototype that

would recommend similar papers to the user. The framework uses the citation network from the AAN dataset. The framework has proven to be more efficient than the other state-of-the-art graph methods concerning recall metrics. The framework provides a recall value of 0.65 for recall@100.

Future studies may include experimentation with other deep learning-based models to improvise the accuracy as well as the relevance of the models. By concentrating on techniques to improve the accuracy of reinforcement learning, further research can find methods to incorporate more deep learning of the node-node interaction using the metadata.

REFERENCES

[1] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 448–456, 2011, doi: 10.1145/2020408.2020480.

[2] J. Sun, J. Ma, Z. Liu, and Y. Miao, "Leveraging content and connections for scientific article recommendation in social computing contexts," *Comput. J.*, vol. 57, no. 9, pp. 1331–1342, 2014, doi: 10.1093/comjnl/bxt086.

[3] H. Liu, Z. Yang, I. Lee, Z. Xu, S. Yu, and F. Xia, "CAR: Incorporating filtered citation relations for scientific article recommendation," *Proc. - 2015 IEEE Int. Conf. Smart City, SmartCity 2015, Held Jointly with 8th IEEE Int. Conf. Soc. Comput. Networking, Soc. 2015, 5th IEEE Int. Conf. Sustain. Comput. Communic*, pp. 513–518, 2015, doi: 10.1109/SmartCity.2015.121.

[4] S. M. McNee *et al.*, "On the recommending of citations for research papers," p. 116, 2002, doi: 10.1145/587095.587096.

[5] C. Bhagavatula, S. Feldman, R. Power, and W. Ammar, "Content-based citation recommendation," *NAACL HLT 2018 - 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 238–251, 2018, doi: 10.18653/v1/n18-1022.

[6] N. J. van Eck and L. Waltman, "Citation-based clustering of publications using CitNetExplorer and VOSviewer," *Scientometrics*, vol. 111, no. 2, pp. 1053–1070, 2017, doi: 10.1007/s11192-017-2300-7.

[7] B. Hu, C. Shi, W. X. Zhao, and P. S. Yu, "Leveraging meta-path based context for top-n recommendation with a neural co-attention model," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 1, pp. 1531–1540, 2018, doi: 10.1145/3219819.3219965.

[8] J. Son and S. B. Kim, "Academic paper recommender system using multilevel simultaneous citation networks," *Decis. Support Syst.*, vol. 105, pp. 24–33, 2018, doi: 10.1016/j.dss.2017.10.011.

[9] H. Liu, X. Kong, X. Bai, W. Wang, T. M. Bekele, and F. Xia, “Context-Based Collaborative Filtering for Citation Recommendation,” *IEEE Access*, vol. 3, pp. 1695–1703, 2015, doi: 10.1109/ACCESS.2015.2481320.

[10] N. Sakib, R. B. Ahmad, and K. Haruna, “A Collaborative Approach Toward Scientific Paper Recommendation Using Citation Context,” *IEEE Access*, vol. 8, pp. 51246–51255, 2020, doi: 10.1109/ACCESS.2020.2980589.

[11] W. Zhang, J. Wang, and W. Feng, “Combining latent factor model with location features for event-based group recommendation,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. Part F1288, pp. 910–918, 2013, doi: 10.1145/2487575.2487646.

[12] Y. Li, M. Yang, and Z. Zhang, “Scientific articles recommendation,” *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 1147–1156, 2013, doi: 10.1145/2505515.2505705.

[13] J. Beel, B. Gipp, S. Langer, and C. Breitinger, “Research-paper recommender systems: a literature survey,” *Int. J. Digit. Libr.*, vol. 17, no. 4, pp. 305–338, 2016, doi: 10.1007/s00799-015-0156-0.

[14] T. Y. Tang and G. McCalla, “A multidimensional paper recommender: Experiments and evaluations,” *IEEE Internet Comput.*, vol. 13, no. 4, pp. 34–41, 2009, doi: 10.1109/MIC.2009.73.

[15] O. Ku¨c¸u¨ktunc¸, E. Saule, K. Kaya, and U.¨mit V. C¸atalyu¨rek, “Direction Awareness in Citation Recommendation,” *DBRank*, vol. i, pp. 3–8, 2012.

[16] C. L. Giles, K. D. Bollacker, and S. Lawrence, “CiteSeer: an automatic citation indexing system,” *Proc. ACM Int. Conf. Digit. Libr.*, pp. 89–98, 1998.

[17] N. Measure, “Co-citation in the Scientific Literature : A New Measure of the Relationship Between Two Documents,” *J. Am. Soc. Inf. Sci.*, vol. 24, no. 4, pp. 265–269, 1973, [Online].

[18] K. D. Bollacker, S. Lawrence, and C. L. Giles, “System for automatic personalized tracking of scientific literature on the Web,” *Proc. ACM Int. Conf. Digit. Libr.*, pp. 105–113, 1999, doi: 10.1145/313238.313270.

[19] R. Dong, L. Tokarchuk, and A. Ma, “Digging Friendship: Paper Recommendation in Social Network,” *Elec.Qmul.Ac.Uk*, pp. 1–7, 2002, [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Digging+Friendship:+Paper+Recommendation+in+Social+Network#0.

[20] X. Tang, X. Wan, and X. Zhang, “Cross-language context-aware citation recommendation in scientific articles,” *SIGIR 2014 - Proc. 37th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 817–826, 2014, doi: 10.1145/2600428.2609564.

[21] D. Hanyurwimfura, L. Bo, V. Havyarimana, D. Njagi, and F. Kagorora, “An Effective Academic Research Papers Recommendation for Non-profiled Users,” *Int. J. Hybrid Inf. Technol.*, vol. 8, no. 3, pp. 255–272, 2015, doi: 10.14257/ijhit.2015.8.3.23.

[22] D. H. Tran *et al.*, “HeteGraph: A Convolutional Framework for Graph Learning in Recommender Systems,” *Proc. Int. Jt. Conf. Neural Networks*, 2020, doi: 10.1109/IJCNN48605.2020.9207078.

[23] C. Jeong, S. Jang, E. Park, and S. Choi, “A context-aware citation recommendation model with BERT and graph convolutional networks,” *Scientometrics*, vol. 124, no. 3, pp. 1907–1922, 2020, doi: 10.1007/s11192-020-03561-y.

[24] W. Waheed, M. Imran, B. Raza, A. K. Malik, and H. A. Khattak, “A Hybrid Approach Toward Research Paper Recommendation Using Centrality Measures and Author Ranking,” *IEEE Access*, vol. 7, no. c, pp. 33145–33158, 2019, doi: 10.1109/ACCESS.2019.2900520.

[25] L. Guo, X. Cai, H. Qin, Y. Guo, F. Li, and G. Tian, “Citation recommendation with a content-sensitive deepwalk based approach,” *IEEE Int. Conf. Data Min. Work. ICDMW*, vol. 2019-Novem, pp. 538–543, 2019, doi: 10.1109/ICDMW.2019.00082.

[26] H. A. Mohamed Hassan, “Personalized research paper recommendation using deep learning,” *UMAP 2017 - Proc. 25th Conf. User Model. Adapt. Pers.*, pp. 327–330, 2017, doi: 10.1145/3079628.3079708.

[27] P. Ganeshkumar and S. Gopalakrishnan, “Systematic reviews and meta-analysis: Understanding the best evidence in primary healthcare,” *J. Fam. Med. Prim. Care*, vol. 2,

no. 1, p. 9, 2013, doi: 10.4103/2249-4863.109934.

[28] A. Grewal, H. Kataria, and I. Dhawan, “Literature search for research planning and identification of research problem,” *Indian J. Anaesth.*, vol. 60, no. 9, pp. 635–639, 2016, doi: 10.4103/0019-5049.190618.

[29] A. Polonioli, “The ethics of scientific recommender systems,” *Scientometrics*, vol. 126, no. 2, pp. 1841–1848, 2021, doi: 10.1007/s11192-020-03766-1.

[30] M. Ceci, N. Japkowicz, J. Liu, G. A. Papadopoulos, and Z. W. Raś, “Preface,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11177 LNAI, pp. v–vi, 2018, doi: 10.1007/978-3-030-01851-1.

[31] L. Yang, Z. Zhang, X. Cai, and T. Dai, “Attention-based personalized encoder-decoder model for local citation recommendation,” *Comput. Intell. Neurosci.*, vol. 2019, 2019, doi: 10.1155/2019/1232581.

[32] N. Nassar, A. Jafar, and Y. Rahhal, “A novel deep multi-criteria collaborative filtering model for recommendation system,” *Knowledge-Based Syst.*, vol. 187, p. 104811, 2020, doi: 10.1016/j.knosys.2019.06.019.

[33] S. Mohammadi, S. Kylasa, G. Kollias, and A. Grama, “Context-Specific Recommendation System for Predicting Similar PubMed Articles,” *IEEE Int. Conf. Data Min. Work. ICDMW*, vol. 0, pp. 1007–1014, 2016, doi: 10.1109/ICDMW.2016.0146.

[34] L. Guo, X. Cai, F. Hao, D. Mu, C. Fang, and L. Yang, “Exploiting Fine-Grained Co-Authorship for Personalized Citation Recommendation,” *IEEE Access*, vol. 5, pp. 12714–12725, 2017, doi: 10.1109/ACCESS.2017.2721934.

[35] J. Rollins, M. McCusker, J. Carlson, and J. Stroll, “Manuscript matcher: A content and bibliometrics-based scholarly journal recommendation system,” *CEUR Workshop Proc.*, vol. 1823, pp. 18–29, 2017.

[36] D. Wang, Y. Liang, D. Xu, X. Feng, and R. Guan, “A content-based recommender system for computer science publications,” *Knowledge-Based Syst.*, vol. 157, pp. 1–9, 2018, doi: 10.1016/j.knosys.2018.05.001.

LIST OF PUBLICATIONS

S.NO.	TITLE	JOURNAL NAME	CATEGORY	PUBLICATION DETAILS
1	Document Classification for Recommender Systems Using Graph Convolutional Networks	Lecture Notes in Networks and Systems	Book Chapter	DOI: https://doi.org/10.1007/978-981-16-4486-3_45
2	Citation Recommendation System Using Deep Reinforcement Learning	Lecture Notes on Data Engineering and Communications Technologies	Book Chapter	DOI: https://doi.org/10.1007/978-981-16-1866-6_31
3	Content Based Scientific Article Recommendation System Using Deep Learning Technique	Lecture Notes in Networks and Systems	Book Chapter	DOI: https://doi.org/10.1007/978-981-16-1395-1_70
4	Similarity Analysis for Citation Recommendation System using Binary Encoded Data	IEEEXplore [IEEE]	Conference Proceedings	DOI: 10.1109/ICECCE49384.2020.9179380
5	Content Based Deep Matrix Factorization Framework for Scientific Article Recommender Systems	Scopus	Accepted as a book chapter in scopus	

LIST OF CONFERENCE PRESENTATIONS

1. International Conference on Data Science Computation and Security (IDSCS 2021)

2. 2nd International Conference on Mobile Computing and Sustainable Informatics (ICMCSI 2021)

Springer
ICMCSI
CERTIFICATE OF PRESENTATION
This certificate is awarded to
Akhil M Nair
for successfully presenting the research paper entitled
An Efficient Citation Recommendation System using Deep Reinforcement Learning
in the 2nd International Conference on
Mobile Computing and Sustainable Informatics (ICMCSI 2021)
organised by Tribhuvan University, Nepal on 29-30, January 2021.
SESSION CHAIR
ORGANIZING SECRETARY
Dr. Jennifer S
CONFERENCE CHAIR
Prof. Dr. Subarna Shakya

3. 5th International Conference on Inventive Systems and Control (ICISC2021)

Springer
JCT
COLLEGE OF ENGINEERING AND TECHNOLOGY
CERTIFICATE OF PRESENTATION
This certificate is awarded to
Akhil M Nair
for successfully presenting the paper entitled
Content based Scientific Article Recommendation System using Deep Learning Technique
in the 5th International Conference on Inventive Systems and Control (ICISC 2021)
organised by JCT College of Engineering and Technology
during 7-8 January, 2021 at Coimbatore, India.
ICISC
SESSION CHAIR
CONFERENCE CHAIR
Dr. K.Geetha
PRINCIPAL
Dr. V. J. Arulkarthick

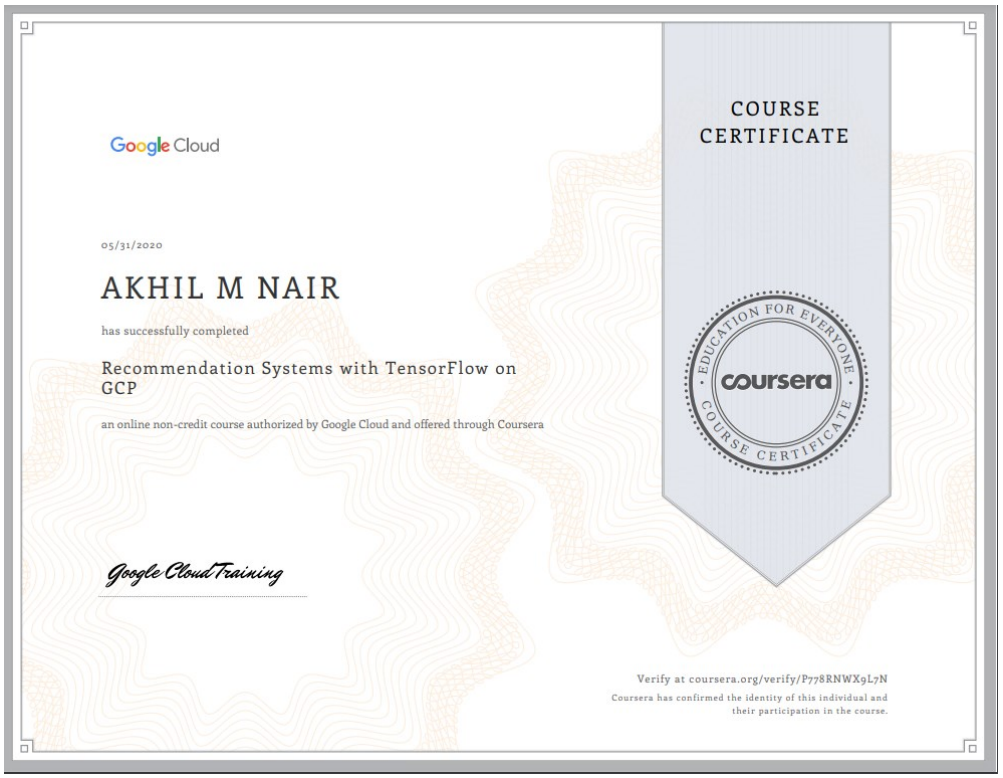
4. 2nd International Conference on Electrical, Communication and Computer Engineering (ICECCE)

5. 3rd International Conference on Data, Engineering and Applications 2021 (IDEA-2k21)

CERTIFICATE
(IDEA - 2K21)
08-09 October, 2021
3rd Springer International Conference
on
Data, Engineering & Applications
(IDEA - 2K21)
Prof/Dr/Mr/Ms Akhil M Nair From Christ University, Bangalore has participated in the 3rd Springer International Conference on Data Engineering & Applications jointly organized by School of Information Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya (The State Technical University of Madhya Pradesh) and Oriental Institute of Science & Technology, Bhopal on 08th and 09th October 2021 and presented a paper titled Content Based Deep Factorization Framework for Scientific Article Recommender System
Dr Rajesh K Shukla
Program Chair
Dr Jitendra Agrawal
Program Chair
Prof Sanjeev Sharma
General Chair
Prof Sunil Kumar
Vice Chancellor, RGPV
Springer
IDEA

CERTIFICATIONS

1. Recommendation Systems with TensorFlow on GCP


Google Cloud
COURSE CERTIFICATE
05/31/2020
AKHIL M NAIR
has successfully completed
Recommendation Systems with TensorFlow on GCP
an online non-credit course authorized by Google Cloud and offered through Coursera
Google Cloud Training
EDUCATION FOR EVERYONE
coursera
COURSE CERTIFICATE
Verify at coursera.org/verify/P778RNWX9L7N
Coursera has confirmed the identity of this individual and their participation in the course.